JOURNAL OF VIROLOGY, Jan. 1990, p. 247–255
0022-538X/90/010247-09$02.00/0

Vol. 64, No. 1

Uukuniemi Virus S RNA Segment: Ambisense Coding Strategy, Packaging of Complementary Strands into Virions, and Homology to Members of the Genus *Phlebovirus*

JAN FREDRIK SIMONS,[1]* ULF HELLMAN,[2] AND RALF F. PETTERSSON[1]

Stockholm Branch, Ludwig Institute for Cancer Research, Box 60202, S-10401 Stockholm,[1] and Uppsala Branch, Box 595, S-75123 Uppsala,[2] Sweden

Received 19 June 1989/Accepted 19 September 1989

We determined the complete nucleotide sequence of the small (S) RNA segment of Uukuniemi virus, the prototype of the *Uukuvirus* genus within the *Bunyaviridae* family. The RNA, which is 1,720 nucleotides long, contains two nonoverlapping open reading frames. The 5′ end of one strand (complementary to the viral strand) encodes the nonstructural protein NS_s (273 residues; molecular weight, 32,019), whereas the 5′ end of the viral-sense strand encodes the nucleocapsid protein N (254 residues; molecular weight, 28,508). Thus, the S RNA uses an ambisense coding strategy previously described for the S segment of two phleboviruses and the arenaviruses. The localization of the N protein within the S RNA sequence was confirmed by amino-terminal sequence analysis of all five possible cyanogen bromide fragments obtained from purified N protein. Northern (RNA) blot analyses with strand-specific probes showed that the N and NS_s proteins are translated from subgenomic mRNAs about 800 and 850 nucleotides long, respectively. These mRNAs are apparently transcribed from full-length S RNAs of opposite polarities. The two mRNA species were also detected in virus-infected cells. Interestingly, highly purified virions contained full-length S RNA copies of both polarities at a ratio of about 10:1. In contrast, virions contained exclusively negative-strand copies of the M RNA segment. The possible significance of these results for viral infection is discussed. The amino acid sequence of the N protein showed 35 and 32% homology (identity) with the N protein of Punta Toro and sandfly fever Sicilian viruses, two members of the *Phlebovirus* genus. The NS_s proteins were much less related (about 15% identity). In addition, the extreme 5′ and 3′ ends of the S RNA, which are complementary to each other, also showed a high degree of conservation with the two phleboviruses. These results indicate that the uukuviruses and phleboviruses are evolutionarily related and suggest that the two genera could be merged into a single genus within the *Bunyaviridae* family.

The *Bunyaviridae* family of arthropod-borne viruses comprises more than 250 different serotypes divided into five genera: *Bunyavirus*, *Hantavirus*, *Nairovirus*, *Phlebovirus*, and *Uukuvirus* (5, 6). All bunyaviruses share the same basic structural characteristics, i.e., four structural proteins and a tripartite segmented RNA genome. All of them also seem to mature in the Golgi complex (20). We are using Uukuniemi (UUK) virus, the prototype of the *Uukuvirus* genus, as a model to study the structure, replication, and assembly of bunyaviruses. The *Uukuvirus* genus has some six members (5) that are transmitted to their vertebrate hosts by ticks.

The genome of UUK virus is split into three single-stranded, circular RNA segments, called L, M, and S, of negative polarity (12, 21, 22, 24, 26). The M segment, which has recently been completely sequenced (27), is 3,231 nucleotides long and encodes a precursor (p110) of the two membrane glycoproteins G1 and G2 that form the surface projections (32) and have an affinity for the Golgi complex (9, 14, 15, 20). The p110 precursor is translated from an mRNA complementary to and probably slightly shorter than the corresponding virion RNA segment (23, 31). The L RNA segment (about 7,500 residues long) encodes the RNA-dependent RNA polymerase L (M_r, about 200,000).

Our previous work has indicated that two small subgenomic 12S mRNA species are transcribed from the S virion RNA segment. In vitro translation of this mRNA fraction gave rise to the nucleocapsid protein N (M_r, about 25,000) and a nonstructural protein, NS_s (M_r, about 30,000), that is found only in infected cells (23, 31). The S segment thus harbors two genes, the organization of which has remained unknown. Analyses of the coding strategy of the S RNA segment of members of the *Bunyavirus* genus has revealed two overlapping genes that encode N and NS_s. The gene for NS_s is completely contained within the N gene, and the proteins are translated in different reading frames (1). For Hantaan virus (*Hantavirus* genus), the S RNA appears to encode only an N protein and no NS_s (30). A totally different strategy has been found for Punta Toro (PT) and sandfly fever Sicilian (SFS) viruses, two phleboviruses. N and NS_s are translated from two smaller, nonoverlapping subgenomic mRNAs of opposite polarities. The N protein is translated from an mRNA corresponding to the 3′ half of the virion RNA, whereas NS_s is made from a virion sense mRNA identical to the 5′ half of the virion RNA (13, 18). This coding strategy, which has also been found for the S RNA segment of members of the *Arenaviridae* family, has been termed ambisense (2, 13).

To analyze the coding strategy of the S RNA of UUK virus, we determined its complete nucleotide sequence from cloned cDNA. The results unambiguously show that UUK virus S RNA also uses an ambisense coding strategy. A surprising finding was that virions seem to package full-length S RNA segments of opposite polarities in about a 10:1 ratio, raising interesting questions about the mechanism of transcription regulation of the S RNA. In contrast, only negative strands of the M segment were packaged into

* Corresponding author.

viruses. A clear sequence homology among the N proteins of UUK, PT, and SFS viruses was found, indicating that uukuviruses and phleboviruses are evolutionarily related.

MATERIALS AND METHODS

Virus and cells. The origin and preparation of stock virus from prototype strain S23 of UUK virus after several successive plaque purifications has been described previously (22). The titer of the virus stock used was 1×10^8 to 2×10^8 PFU/ml. Cells were infected with a multiplicity of about 10 PFU per cell. The virus was grown in BHK21 cells, clone 13, or chicken embryo cells, in 1-liter roller bottles or 250-ml plastic flasks in Eagle minimal essential medium supplemented with 0.2% bovine serum albumin, 20 mM HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid), pH 7.4, and antibiotics (22).

Isolation of viral RNA and virus-specific mRNA. The S RNA segment was isolated from purified virus (about 5.8 mg of protein) as described previously (27). The purification involved isolation of the ribonucleoproteins (RNPs) from Triton X-100-solubilized virus on a CsCl gradient, followed by separation of the L, M, and S RNA species on a 15 to 30% (wt/wt) sucrose gradient after treatment with 1% sodium dodecyl sulfate (SDS). The three RNA peak fractions were pooled individually and concentrated by ethanol precipitation. The final yield of S RNA was about 10 μg.

To isolate virus-specific mRNAs, BHK21 cells were infected with a multiplicity of about 10 PFU per cell. Dactinomycin (2 μg/ml) was added at 16 h postinfection. At 30 min later, 50 μCi of [^{3}H]uridine (Amersham Corp.; specific activity, 25 to 30 Ci/mmol) per ml was added. At 22 h postinfection, the cells were washed with phosphate-buffered saline, trypsinized, and collected by centrifugation. The cells were then solubilized with 1% Triton X-100 for 10 min on ice in 10 mM Tris hydrochloride (pH 7.4)–150 mM NaCl–1.5 mM $MgCl_2$ (iso-B buffer) containing 40 μg of polyvinyl sulfate per ml as an RNase inhibitor. Nuclei were removed by low-speed centrifugation, and the supernatant was mixed with CsCl in 10 mM Tris hydrochloride (pH 7.4)–100 mM NaCl–1 mM EDTA (TNE buffer) to give an initial density of 1.31 g/ml. The mRNAs were then separated from the RNPs by centrifugation in an SW50.1 rotor at 45,000 rpm for 16 h at 4°C. The mRNA from the pellet was suspended in TNE buffer containing 1% SDS and further fractionated on a 15 to 30% sucrose gradient as described previously (31). Peak fractions were pooled, and the RNA species were concentrated by ethanol precipitation.

Cloning and sequence determination. Synthesis of cDNA was carried out essentially as described by Gubler and Hoffman (10) and Rönnholm and Pettersson (27). For first-strand synthesis, purified S RNA (about 2 μg per reaction) was used as the template and a synthetic oligonucleotide complementary to the first 10 3'-terminal nucleotides of the M virion RNA (27) was used as a primer. This oligonucleotide, which has the sequence 5' CGAGCTGCAGACA CAAAGAC 3', also contains a cleavage site for *PstI* and four terminal protecting nucleotides upstream from the complementary sequence. Before synthesis, the RNA was denatured in 2.5 mM methyl mercury hydroxide (3) for 4 min at room temperature, followed by inactivation of the denaturant with 15 mM (final concentration) β-mercaptoethanol. The double-stranded cDNA was blunt ended with T4 DNA polymerase (Boehringer Mannheim Biochemicals), ligated into the similarly blunt-ended *PstI* site of pBR322, and cloned into *Escherichia coli* HB101. Recombinant plasmids were isolated from cleared lysates, fractionated on 1% agarose gels, blotted onto GeneScreen filters (Dupont, NEN Research Products), and analyzed by hybridization with purified ^{32}P-labeled S virion RNA prepared as described previously (21). One clone, designated pUS1037, containing a complete S RNA insert was used for further analyses.

Sequence determination was performed by a modification of the dideoxy-chain termination method (29) using the double-stranded cloned cDNA as the template and Sequenase (United States Biochemical Corp.). The sequence was determined from both strands by using a total of 17 synthetic oligonucleotides (18-mers) spaced at an interval of about every 250 nucleotides along the insert.

Northern (RNA) and slot blot analyses. Viral RNA was prepared from virus grown in chicken embryo cells and purified on a sucrose gradient as described previously (27). RNA was solubilized in TNE buffer containing 2% SDS, followed by two extractions with phenol-chloroform (1:1) and ethanol precipitation.

RNA to be analyzed was resolved on a 1% denaturing agarose gel containing 7% formaldehyde in 20 mM morpholinepropanesulfonic acid (MOPS; pH 7.0)–50 mM sodium acetate–1 mM EDTA (MOPS buffer). Before loading, the RNA was denatured for 3 min at 70°C in 50% formamide–7% formaldehyde in MOPS buffer. The gel was run at 80 V for about 3 h, and the RNA was blotted onto Hybond-N filters (Amersham Corp.). The filters were hybridized to 5'-end ^{32}P-labeled oligonucleotide probes. The probes used were R84 and R85, which are identical to nucleotides 1458 to 1502 and 225 to 269 (see the lower strand in Fig. 1), respectively, and R86 and R87, which have a polarity complementary to that of the viral-sense RNA at nucleotides 1416 to 1461 and 132 to 177 (see the upper strand in Fig. 1). For M RNA, probe R88, which has the same sequence as the viral-sense RNA at positions 1533 to 1487 (27), and R89, which is complementary to positions 1438 to 1482 of the viral-sense RNA, were used.

For slot blot analyses, virion RNA was diluted in a twofold series and applied to Hybond-N membranes (Amersham) by the procedure recommended by the manufacturer. The filters were hybridized to [^{32}P]UTP-labeled RNAs transcribed in vitro from two recombinant plasmids containing S RNA-specific inserts. The insert in pGEM-3N yielded a 960-residue-long transcript complementary to the virion RNA, whereas pGEM-3NS was transcribed into a 1,060 residue-long RNA that hybridized to the virus-complementary strand. Both probes were shown to detect the two strands equally efficiently by hybridization to serial dilutions of polymerase chain reaction-amplified double-stranded DNA corresponding to full-length S RNA.

Primer extension analysis. On the basis of the complete nucleotide sequence, two oligonucleotides (R82[5' ACA CAAAGACCCTCCAACATTAAGCA 3'] and R83[5' ACA CAAAGACGTCCAACTTAGCTATCG 3']) complementary to and specific for the extreme 3' ends of the full-length S RNAs of opposite polarities were synthesized and used in primer extension experiments. The conditions were the same as those used for first-strand cDNA synthesis (see above). The ^{32}P-labeled products were analyzed by electrophoresis in a 1% alkaline agarose gel, followed by autoradiography.

Purification and cyanogen bromide cleavage of N protein; amino-terminal sequencing of isolated fragments. N protein was isolated from purified RNPs after treatment with 2% SDS, followed by sucrose gradient centrifugation as described for isolation of virion RNAs (see above). N protein

was recovered by pooling the top fractions of the sucrose gradient, followed by dialysis against distilled water and lyophilization. SDS was removed by ion pair extraction by the method of Henderson et al. (11). The protein was further purified on a Superose 12 fast-protein liquid chromatography column (Pharmacia-LKB, Uppsala, Sweden) in 70% formic acid. N protein was then cleaved by treatment with 5 mg of cyanogen bromide (CNBr) per ml in 70% formic acid at room temperature in the dark and in a nitrogen atmosphere. After 17 h, the sample was diluted 10-fold with water and lyophilized.

The lyophilized sample was subsequently dissolved in 200 μl of ultrapure 6 M guanidine hydrochloride, and the CNBr-generated fragments were separated by reversed-phase high-pressure liquid chromatography on a narrow-bore (2.1 by 30 mm) Brownlee Aquapore C4 column with a linear gradient of acetonitrile in 0.1% (vol/vol) trifluoroacetic acid. The isolated fragments obtained were subjected to amino-terminal sequencing by an Applied Biosystems 470A sequencer equipped with an on-line 120A phenylthiohydantoin amino acid analyzer.

Computer analyses. Computer analyses were performed by using the programs of the University of Wisconsin Genetics Computer Group and executed on a VAX 11/750 computer. The hydropathy plot was obtained by the method of Kyte and Doolittle (16). The amino acid comparisons were done by using the GAP program, which is based on the method of Needleman and Wunsch (19). The dot plots were derived by the COMPARE program (17) in conjunction with a comparison table scoring 1.5 for identity and 0 for conserved changes and nonidentity. The comparison window was 30, and the stringency was 12, making a dot for every window of 30 residues with eight or more identical amino acids.

RESULTS

Cloning and sequencing of S cDNA. The first 10 to 12 5′- and 3′-terminal nucleotides of the L, M, and S RNA segments of bunyaviruses have been found to be complementary to each other and conserved in the RNAs of each virus and even in members of the same genus. We have previously determined the 5′- and 3′-end sequences of the UUK virus M RNA segment (27). Assuming that these sequences are also conserved in the S RNA, we synthesized a 10-residue-long primer complementary to the 3′ end of the M RNA. Using this primer and S RNA as a template, we obtained a single-stranded cDNA product of approximately 1,700 bases (data not shown). This is close to the size of the S RNA previously determined by a variety of methods (21). Double-stranded cDNA was therefore synthesized, ligated into pBR322, and cloned in *E. coli*. One clone (pUS1037) specifically hybridized to ^{32}P-labeled S virion RNA and appeared to contain a full-length insert. This insert was completely sequenced from both strands by using synthetic oligonucleotide primers. The sequence comprising 1,720 nucleotides is shown in Fig. 1. Because of reasons made apparent below, the sequences of both complementary strands are shown. At the 3′ end of one of the strands (Fig. 1, lower strand) the sequence 3′ UGUGUUUCUG. . ., which is complementary to the primer used, was identified. At the 5′ end of the same strand, the sequence 5′ ACACAAAGAC. . . was found. This sequence is perfectly complementary to the 3′ end. Both sequences are also identical to the ends of the M RNA segment (see Table 1). We therefore conclude that the sequence of the insert in pUS1037 represents full-length S RNA.

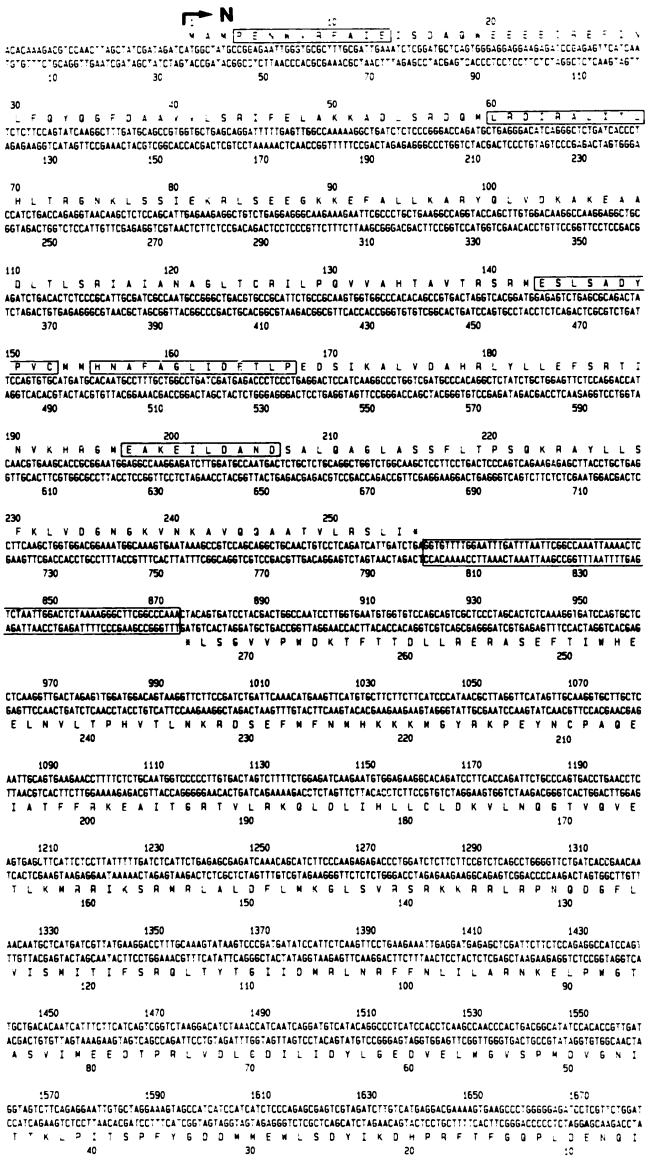

FIG. 1. Nucleotide sequence of UUK virus S RNA segment and deduced amino acid sequences of N and NS$_S$ proteins. The sequences (1,720 nucleotides) of both complementary strands are shown as DNA. The sequences have been numbered from the 5′ end of the vcRNA (see the text). Note that the N and NS$_S$ proteins are encoded within the 5′ halves of the two strands of opposite directions. The amino acids have been numbered from the amino termini. Amino acid sequences determined by sequencing of CNBr fragments derived from the N protein are boxed. The intergenic noncoding region between nucleotides 800 and 873 is also boxed.

Characteristics of the S RNA sequence. The sequence in Fig. 1 (1,720 nucleotides) corresponds to a molecular weight of 590,000, which is close to the value obtained previously by several methods (21). The base composition of the RNA is 25.1% C, 24.9% A, 24.9% G, and 25.1% U.

The 5′ and 3′ ends are completely complementary for the first 10 nucleotides and partially complementary for some 20

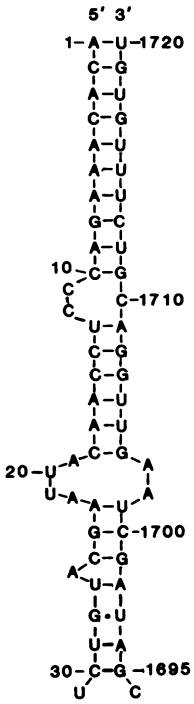

FIG. 2. Computer prediction of the base-paired structure formed between the inverted complementary sequence at the 5′ and 3′ ends of the vRNA.

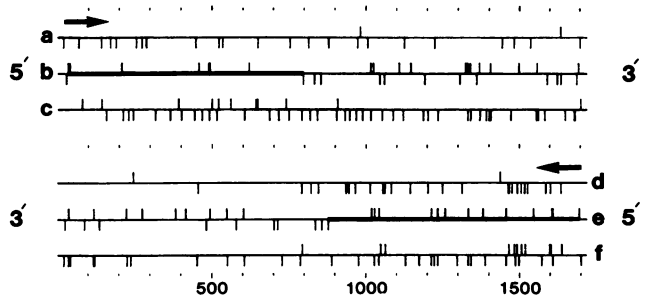

FIG. 3. Distribution of translation termination codons (vertical bars below the lines) and initiation (AUG) codons (vertical bars above the lines) in all six reading frames of vcRNA (frames a to c) and vRNA (frames d to f). Arrows indicate directions of translation. The ORFs of proteins N (frame b) and NS_s (frame e) are indicated as thick lines. Numbers at the bottom indicate nucleotides.

additional nucleotides (Fig. 2). Thereafter, the complementarity is low. The calculated free energy of the stem structure shown in Fig. 2 is −27.8 kcal/mol by the method of Freier et al. (8) or −43.9 kcal/mmol by the method of Salser (28). This stem structure is very similar in size and calculated stability to the one present in M RNA (27) and offers an explanation for the circularity of the S RNP and RNA observed by electron microscopy (12, 24).

The 5′-end sequences of the S RNA of UUK virus is identical to that of the 5′ end of PT and SFS S RNAs for the first 12 nucleotides (Table 1). A similar, although less striking, homology is also apparent at the 3′ ends of these RNAs. Table 1 also shows a comparison of the 3′ and 5′ ends of the S and M RNAs of UUK virus.

Two large ORFs in S RNA. Computer analysis revealed two large, nonoverlapping open reading frames (ORFs) (Fig. 3). One of the ORFs (reading frame b) is located between an AUG codon at position 35 to 37 (Fig. 1) and a UGA codon at position 797 to 799. This ORF corresponds to a product 254 amino acids (molecular weight, 28,508) long. The other ORF (reading frame e) is located within the 5′ end of the complementary strand between an AUG at position 1696 to 1694 and a UAG codon at position 876 to 874. The deduced size of this product is 273 residues (molecular weight, 32,019). The sizes of these putative products are slightly larger than those previously determined for the N (M_r, 25,000) and NS_s (M_r, 30,000) (31) proteins, respectively.

To identify which of the ORFs corresponds to the N protein, we determined the partial amino-terminal sequences of purified CNBr fragments derived from the N protein. The amino-terminal sequences of five such fragments were identified in the 254-residue ORF and are boxed in Fig. 1. These results unambiguously identified this ORF as the one that codes for the N protein. In addition to the initiating amino-terminal methionine codon, the N protein contains six methionine residues, two of which are located next to each other (residues 153 and 154). Since CNBr cleaves after methionine residues, this means that amino-terminal sequence information was obtained from all large CNBr fragments. Initially, we attempted to sequence the amino terminus of the intact N protein but were unsuccessful in obtaining any sequence. This suggested that the amino terminus of the N protein and, consequently, that of the amino-terminal CNBr fragment are blocked. At the amino terminus of the deduced N protein sequence, there are two in-frame AUG codons separated by a codon for alanine. This raised the question of which of the methionine codons initiates the N protein. Since the amino-terminal sequence of the CNBr fragment immediately downstream from the second methionine was readily obtained and the amino terminus of the intact protein was blocked, presumably at the alanine

TABLE 1. Comparison of the inverted complementary 5′- and 3′-terminal sequences of UUK, PT, and SFS virus S RNAs[a]

Virus (RNA)	5′ sequence	3′ sequence
	1 10 20 30	30 20 10 1
PT (S)	5′ ACACAAAGACCCCCUGAAUUUUUUCGUCAU	UAUUAAAUAAAUUUUCAGGGAGCUAUGUGU 3′
UUK (S)	ACACAAAGACCCUCCAACAUUAAGCAUGUC	UAUCGAUAGCUAAGUUGGACGUCUUUGUGU
SFS (S)	ACACAAAGACCCCCUAGUCAAUGAUCAACA	CUCACUCAGAUUAACUAGGGACCUUUGUGU
UUK (S)	5′ ACACAAAGACCCUCCAACAUUAAGCAUGUC	UAUCGAUAGCUAAGUUGGACGUCUUUGUGU 3′
UUK (M)	ACACAAAGACACGGCUACAUGGAACAACAA	AUGUCCUUACCAUGUUAGCCGUCUUUGUGU

[a] Asterisks indicate identical nucleotides in the sequences compared.

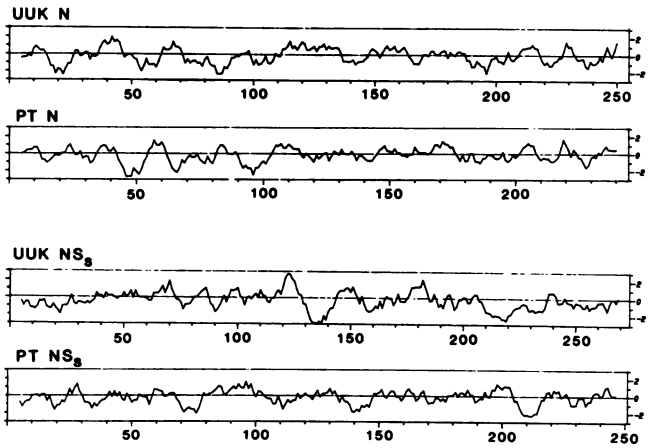

FIG. 4. Hydropathy plots of UUK and PT (13) virus N and NS_s proteins. Hydrophilic regions are above and hydrophobic regions are below the middle lines. Numbers indicate amino acid residues. Note that the N and NS_s proteins have different numbers of residues in the two viruses.

residue, we conclude that the first methionine codon most likely initiates the N protein.

No amino acid sequence information derived from the NS_s protein itself is available. The identification of the second large ORF as the one that corresponds to the NS_s protein, therefore, remains tentative. However, no other ORF large enough to encode the NS_s protein was found in any of the six reading frames on either strand (Fig. 3).

The deduced N and NS_s sequences revealed no unusual features (Fig. 1). Likewise, the hydropathy profiles of N and NS_s (Fig. 4) revealed no striking characteristics, such as long hydrophobic or hydrophilic regions.

The two ORFs have 5′ noncoding regions of 34 (upstream of the N ORF) and 25 (upstream of the NS_s ORF) nucleotides, respectively. Between the two ORFs, there is an intergenic region of 74 nucleotides, which is A+U rich (62%). Notably, there are short runs of U and A bases within this region.

Confirmation of ambisense coding strategy by hybridization. The results described above indicated that the N and NS_s proteins are expressed from two nonoverlapping ORFs by an ambisense coding strategy from RNAs of opposite polarities. To confirm this, we analyzed both virion RNA and mRNAs isolated from infected cells by Northern blotting by using two complementary pairs of oligonucleotides as probes. One pair (R85 and R87) mapped to the N protein region, whereas the other pair (R84 and R86) mapped to the NS_s region. R84 and R85 were of the same polarity and were complementary to the RNA strand in which the ORF for the N protein was located at the 5′ end. R86 and R87 had the same polarity and should hybridize to the RNA strand that coded for the NS_s protein at its 5′ end (see Fig. 8 for clarification). Figure 5A shows the results of the hybridizations. R86 and R87 both hybridized strongly to a full-length S RNA species. R86, in addition, detected a smaller RNA about 800 nucleotides long, whereas R87 failed to detect this RNA species. This indicated that virus particles contain an S RNA with a polarity corresponding to that of the coding sequence of the NS_s protein, and infected cells contain a subgenomic S mRNA with the same polarity as this RNA and, thus, the capability of coding for NS_s.

Hybridization with R84 and R85, which have a polarity opposite to that of R86 and R87, gave an unexpected result

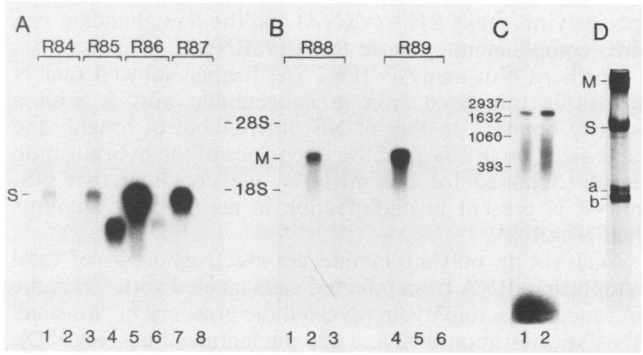

FIG. 5. Northern blot analyses of viral RNAs and mRNAs with strand-specific oligonucleotide probes. (A) Total RNA was isolated from virus particles, or S mRNA was isolated from virus-infected cells (at 22 h postinfection), denatured, and fractionated on a 1% agarose gel containing formaldehyde. An equal amount (about 100 ng) of vRNA was loaded onto lanes 1, 3, 5, and 7, or S mRNA was loaded onto lanes 2, 4, 6, and 8. After transfer to nylon filters, the RNAs were hybridized to four synthetic 5′-end ^{32}P-labeled oligonucleotides. R84 and R86 mapped to the NS_s gene region and were of opposite polarities, whereas R85 and R87 mapped to the N region and were also of opposite polarities (see the legend to Fig. 8 for clarification). The positions of full-length S RNA and 28S and 18S RNAs are indicated. (B) Hybridization of two M RNA segment-specific oligonucleotide probes, R88 and R89, to viral RNAs. R88 was of the same polarity as virion-sense RNA, whereas R89 had mRNA-sense polarity. Total virion RNA (lanes 1 and 4), purified M mRNA isolated from infected cells (lanes 2 and 5), and purified S subgenomic 12S mRNAs (lanes 3 and 6) were run on a denaturing agarose gel as for panel A. (C) Primer extension analyses performed by using two strand-specific oligonucleotide primers (see Materials and Methods) complementary to the 3′ ends of vcRNA (lane 1) and vRNA (lane 2). Equal amounts of a virion S RNA segment (about 100 ng) were used as templates in both lanes. cDNA synthesis was carried out as described in Materials and Methods by using [^{32}P]dCTP as the label. The single-stranded cDNA products were analyzed on an alkaline 1% agarose gel. The positions and sizes of single-stranded DNA marker fragments are indicated to the left. (D) Analysis by polyacrylamide gel electrophoresis of virus-specific total cytoplasmic RNA. Virus-specific RNA was labeled with ^{32}P at 4 to 6 h postinfection in 2 μg of dactinomycin per ml. Following phenol extraction and ethanol precipitation, the RNA was denatured with 1 M glyoxal and 50% dimethyl sulfoxide at 50°C for 1 h and analyzed on a 2.5 to 5% gradient polyacrylamide gel. The positions of the M and S vRNA species and the two subgenomic S RNA-derived mRNAs (bands a and b) are indicated. L RNA did not enter the gel because of its large size.

in that both probes hybridized weakly to a full-length S RNA species. In addition, R85 strongly hybridized to a subgenomic mRNA species from infected cells with a size of about 800 nucleotides. R84 failed to hybridize to this species. These results indicate that virus particles also contain, in addition to the predominant S RNA species detected with R86 and R87, small amounts (about 10%) of full-length complementary S RNA. Because the different probes hybridized to the RNA species with slightly different efficiencies, as revealed by cross-hybridization of the filters with the different probes, it was not possible to estimate accurately on the basis of these blots the amounts of the respective RNAs packaged into virions. Serial dilutions were therefore subjected to slot blot hybridization using [^{32}P]UTP-labeled strand-specific riboprobes (see Materials and Methods). Quantitation of the hybridization signals confirmed that in this particular RNA preparation, the two strands were present at a ratio of about 1:10 (data not shown). For clarity of discussion, we call the abundant RNA

species viral-sense RNA (vRNA) and the less abundant one virus complementary-sense RNA (vcRNA).

Northern blot analysis (Fig. 5A) further showed that N protein is translated from a subgenomic mRNA with a polarity opposite to that of NS_s mRNA but of roughly the same size. On the basis of the intensities of the hybridization signals obtained for the mRNAs, we conclude that NS_s mRNA is present in midinfection in much lower amounts than N mRNA.

Analysis by polyacrylamide gel electrophoresis of total cytoplasmic RNA from infected cells labeled with ^{32}P in the presence of dactinomycin revealed the presence of two small RNA species about 800 and 850 nucleotides long (Fig. 5D). The larger one (band a in Fig. 5D) was less abundant than the faster migrating one (band b). A pool of these two RNA species has previously been shown by in vitro translation to direct the synthesis of the NS_s and N proteins (31).

The presence in virus particles of full-length RNAs of both polarities was further confirmed by cDNA synthesis using two strand-specific synthetic oligonucleotide primers, R82 and R83, complementary to the 3′ ends of the two strands, respectively. Both oligonucleotides were able to prime the synthesis of full-length cDNAs (Fig. 5C). To find out whether both strands of the M RNA segment are also packaged into virions, Northern blot analysis of purified virion RNA was carried out by using two strand-specific oligonucleotides of opposite polarities (R88 and R89). R89 readily hybridized to full-length M RNA (Fig. 5B, lane 4) but failed to detect isolated M mRNA (lane 5) or S mRNAs (lane 6). In contrast, R88 did not hybridize to virion RNA (lane 1) or S mRNAs (lane 3) but readily detected M mRNA from infected cells. Thus, we found no indication that the M RNA of mRNA sense is packaged into virions.

Homology among the sequences of the N proteins of UUK, PT, and SFS viruses. Previous results indicated that there is low but significant homology among the glycoproteins of the UUK, Rift Valley Fever, and PT viruses (27). Dot matrix analyses showed that the N proteins of the UUK, PT, and SFS viruses are clearly related, as evidenced by the diagonals in the three panels of Fig. 6. In contrast, similar comparisons of the NS_s proteins revealed no apparent homologies (data not shown). In Fig. 7, the amino acid sequences of the N proteins of the three viruses are aligned for maximum homology. If both identical amino acids (thick bars) and conserved changes (thin bars) are taken into account, the level of homology between the N proteins of UUK and PT is about 56% and that between the UUK and SFS N proteins is about 60%. If only identity is scored, then the levels of homology are about 35 and 32%, respectively. The corresponding figures for the NS_s proteins are 39 and 35% (homologies including conserved changes) and 15 and 15% (identity), indicating that they are very distantly related. Despite an apparently low degree of homology on the primary sequence level, proteins may still exhibit similar hydropathy profiles. The profiles of the N proteins of the UUK and PT viruses are rather similar, in conformity with the primary sequence comparison (Fig. 4, upper panels). In contrast, the NS_s profiles (lower panels) display no apparent similarities, further strengthening the conclusion that the NS_s proteins are not closely related.

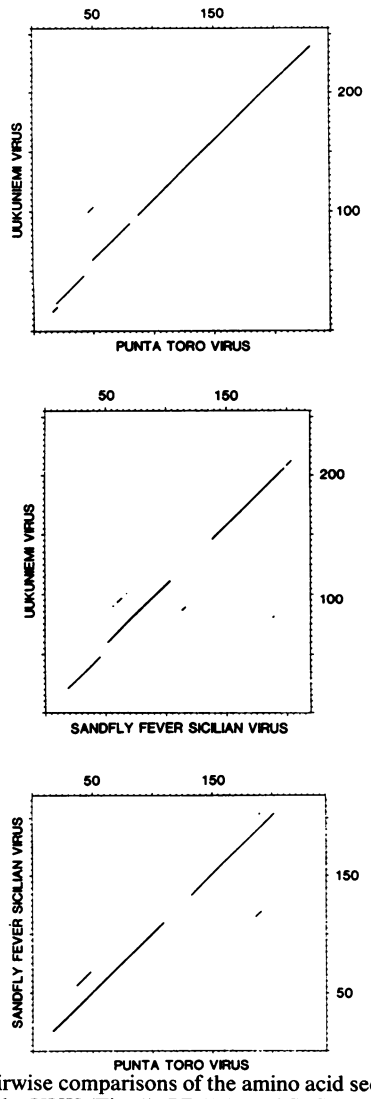

FIG. 6. Pairwise comparisons of the amino acid sequences of the N proteins of the UUK (Fig. 1), PT (13), and SFS (18) viruses by dot matrix analyses. The comparisons were made by using the parameters described in Materials and Methods.

DISCUSSION

In this report, we present data showing that the small S RNA segment of UUK virus expresses two proteins, nucleocapsid protein N and the previously described (31) nonstructural protein NS_s, from two subgenomic mRNAs of opposite polarities. Sequence determination revealed that the N and NS_s proteins are encoded by ORFs present in the 5′ halves of S RNAs of opposite polarities. This coding strategy has been called ambisense and was first described for the S RNA of PT virus (13) and recently also for that of SFS virus (18), both members of the *Phlebovirus* genus within the *Bunyaviridae* family. A similar strategy has also been found for the S RNA segment of arenaviruses (e.g., Pichinde virus) (2). In the latter case, the S RNA encodes a precursor for the two glycoproteins and the nucleocapsid protein. As discussed below, we found that full-length S RNAs of both polarities are packaged into UUK virions at a ratio of about 10:1. For clarity and in agreement with the results obtained for the PT and SFS viruses, we have called the dominant species vRNA and the less abundant species vcRNA, although both strands are packaged into virions. Figure 8 summarizes our view of the coding strategy for the UUK virus S RNA segment.

The N protein was found to be encoded by an ORF located

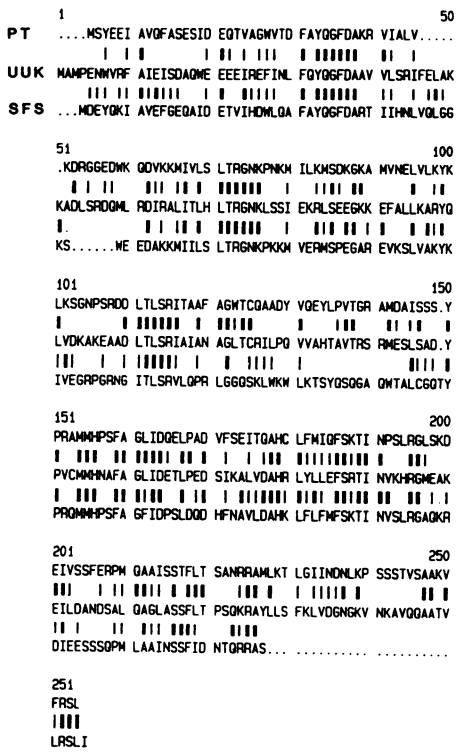

FIG. 7. Comparison of the amino acid sequences of the N proteins of the UUK (Fig. 1), PT (13), and SFS (18) viruses. The sequences were aligned for maximum homology. Thick bars indicate identical residues, and thin bars indicate conserved amino acid changes.

in the 5′ half of vcRNA, since five CNBr fragments for which partial amino-terminal amino acid sequences were obtained were found in this ORF. NS_s must be encoded by the other large ORF found in the 5′ half of vRNA, since no other ORF large enough to specify this protein was found in either strand (Fig. 3). Since amino acid sequence data derived from the NS_s protein is lacking, formal proof is still needed to substantiate this conclusion. Preliminary in vitro transcription-translation of this region has yielded a product with the size of NS_s (J. F. Simons, unpublished data).

We have previously shown that N and NS_s could be translated in vitro from a sucrose gradient peak fraction containing two small 12S virion-specific mRNAs (23, 31)

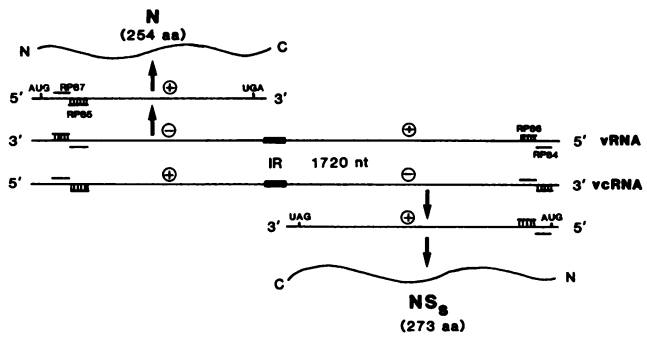

FIG. 8. Schematic representation of the ambisense coding strategy of the UUK virus RNA segment. The localization of probes R84 to R87 used for Fig. 5 and their hybridization specificities are also indicated. nt, Nucleotides; aa, amino acids.

isolated from infected cells. NS_s is made in infected cells in much lower quantities than N (31). As shown here, the larger subgenomic mRNA (about 850 nucleotides) is less abundant than the smaller one (about 800 nucleotides). Thus, it is likely that the larger mRNA, on the basis of both size and abundance, encodes NS_s, whereas the smaller mRNA encodes the N protein. Hybridization experiments with strand-specific oligonucleotides showed that these two mRNAs have opposite polarities, further confirming the ambisense coding strategy. Subgenomic mRNAs have also been demonstrated for the PT (13), SFS (18), and Pichinde (2) viruses.

An unexpected finding was that UUK virus particles contained full-length S vRNA and vcRNA strands at a ratio of approximately 10:1. Previous fingerprint analyses by two-dimensional gel electrophoresis of RNase T_1 oligonucleotides derived from viral S RNA revealed a background of submolar amounts of oligonucleotides, the origin of which could not be determined (21). In light of the present results, it is likely that these oligonucleotides were derived from the vcRNA strand. Packaging of the vcRNA strands has not been reported for the PT and SFS viruses (13, 18), but their presence might have been overlooked, since double-stranded probes were used in the hybridization experiments. Interestingly, packaging of the UUK virus vcRNA strand was specific for the S RNA segment, since only negative strands of the M RNA segment were found in the same virus preparation. This raises the question of the role, if any, of the S vcRNA strands early in the infection. The function of NS_s is not known, although a role as a component in the replication-transcription machinery is one possibility. NS_s has not been found in highly purified UUK virus preparations. Thus, if NS_s is required in replication, there is an apparent paradox in that the mRNA for NS_s can be transcribed only after the viral RNA has first been replicated into the vcRNA. Our findings reported here suggest that this is not necessary if virus particles contain both vRNA and vcRNA strands. In such a case, both mRNAs could be produced through primary transcription carried out by the virion RNA polymerase (26). If this were true, then one would expect the two strands to be present in each infectious particle, i.e., they should be present at an equimolar ratio in virus preparations. If every particle contained only one copy of each of the three RNA segments, then our results, which indicated a ratio of roughly 10:1 of vRNA to vcRNA, would mean that only every tenth particle would contain the vcRNA strand. In UUK virus preparations, the ratio of L to M to S RNAs varies from one preparation to another (R. F. Pettersson, unpublished data) and has never been found to be equimolar. Instead, the average molar ratio is about 1:4:2 (21, 22). This would mean that, as a minimum, the vcRNA would be present, on average, in every fifth particle. Since it is possible that virus particles could package several copies of each RNA segment, an even higher fraction of particles could contain S vcRNA. If vcRNA is indeed needed to initiate infection, then only particles containing both vRNA and vcRNA would be infectious. This would mean that the amount of S vcRNA packaged is the limiting factor for production of infectious virions. It is also fully possible that the presence of vcRNA in virions simply represents mispackaging and that vcRNA is not needed to initiate infection. It should be noted that in La Crosse virus (a bunyavirus), S vcRNA has been detected in virions produced in insect cells, whereas it was not found in virions produced in mammalian cells (25). In this case, however, a different coding strategy is used. The N and NS_s proteins are translated from the same

mRNA strand from overlapping ORFs by using two different frames (1).

Between the two ORFs that encode N and NS_s, there is a noncoding intergenic region (IR) of only 74 nucleotides. This is much shorter than the corresponding regions in PT (360 nucleotides) and SFS (250 nucleotides) (13, 18). In the S RNA of UUK virus, the IR is A+U rich and computer analysis shows that it can be formed into a branched stem-loop structure with a calculated free energy of only about −11.8 kcal/mol (8; data not shown). In PT virus S RNA, the IR is also A+U rich and contains a long inverted complementary region that can be formed into a stem-loop structure. Transcription termination of the N and NS_s mRNAs occurs close to the terminal loop of the stem structure (7). In contrast, the IR in the S vRNA of SFS virus is very C rich but cannot be folded into a clear base-paired hairpin structure. Thus, it appears that the IRs in these three viruses are very dissimilar in structure. The termination sites for the UUK virus mRNAs have yet to be defined.

As for the M RNA segment of UUK virus (27), the 5′ and 3′ ends of the S RNA are inverted complementary to each other and can form stable panhandle structures. This confirms conclusions drawn on the basis of electron microscopy studies showing that both M and S RNPs (24) and RNAs (12) have a circular configuration that can be dissociated into linear molecules by denaturation. Complementary ends of the M and S RNA segments have also been found for all bunyaviruses for which sequence information has been obtained. The significance of these base-paired panhandle structures is unknown, although an important role in replication and/or packaging is likely (4, 12).

Phleboviruses and uukuviruses have been classified into separate genera, since extensive analyses have shown them to be serologically unrelated. Recently, however, weak cross-reactivity has been observed between members of these two genera (C. H. Calisher, personal communication). It was therefore interesting to find a rather high degree of sequence homology among the N proteins of the UUK, PT, and SFS viruses (about 32 to 35% identity and 55 to 60% similarity). A much lower (about 15 to 25%) but still significant level of homology was recently also found between glycoproteins G1 and G2 of UUK virus, and those of the PT and Rift Valley fever viruses (27). Furthermore, in this study, we found a striking homology between the 5′- and 3′-end nucleotide sequences of the M and S RNAs of these viruses. In contrast, NS_s of UUK virus displayed a very low degree of homology to the PT and SFS virus NS_ss, whereas the NS_ss of the PT and SFS viruses were reported to share about 25% identical amino acids. To summarize, the available data indicate that members of the *Phlebovirus* and *Uukuvirus* genera are closely related because (i) they share the same ambisense coding strategy for the S RNA segment, (ii) they have identical 5′- and 3′-terminal nucleotide sequences, (iii) they display low but significant homology between the glycoproteins, and (iv) the N proteins show a high degree of homology. In light of these new facts, reclassification of the *Phlebovirus* and *Uukuvirus* genera into one single genus seems justified.

ACKNOWLEDGMENTS

We thank Ulf Eriksson for help with purification and CNBr cleavage of the N protein and Christer Wernstedt at the Uppsala Branch of the Ludwig Institute for Cancer Research for sequence determinations of the CNBr fragments. We are also grateful to Maj-Greth Nilsson for technical advice and assistance and Robert Persson for valuable discussions.

LITERATURE CITED

1. **Akashi, H., and D. H. L. Bishop.** 1983. Comparison of the sequences and coding of La Crosse and snowshoe hare bunyavirus S RNA species. J. Virol. **45**:1155–1158.
2. **Auperin, D. D., V. Romanowski, M. Galinski, and D. H. L. Bishop.** 1984. Sequencing studies of Pichinde arenavirus S RNA indicate a novel coding strategy, an ambisense viral S RNA. J. Virol. **52**:897–904.
3. **Bailey, J. M., and N. Davidson.** 1976. Methylmercury as a reversible denaturing agent for agarose gel electrophoresis. Anal. Biochem. **70**:75–85.
4. **Baltimore, D., R. F. Pettersson, J. B. Flanegan, M. J. Hewlett, J. K. Rose, and V. Ambros.** 1978. New structures in viral RNA: non-covalent circles and covalently linked proteins. Perspect. Virol. **10**:1–12.
5. **Bishop, D. H. L.** 1985. Replication of arenaviruses and bunyaviruses, p. 1083–1110. *In* B. N. Fields et al. (ed.), Virology. Raven Press, New York.
6. **Bishop, D. H. L., C. H. Calisher, J. Casals, M. P. Chumakov, S. Y. Gaidamovich, C. Hannoun, D. K. Lvov, I. D. Marshall, N. Oker-Blom, R. F. Pettersson, J. S. Porterfield, P. K. Russell, R. E. Shope, and E. G. Westaway.** 1980. Bunyaviridae. Intervirology **14**:125–143.
7. **Emery, V. C., and D. H. L. Bishop.** 1987. Characterization of Punta Toro S mRNA species and identification of an inverted complementary sequence in the intergenic region of Punta Toro phlebovirus ambisense S RNA that is involved in mRNA transcription termination. Virology **156**:1–11.
8. **Freier, S. M., R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson, and H. Turner.** 1986. Improved free-energy parameters for predictions of RNA duplex stability. Proc. Natl. Acad. Sci. USA **83**:9373–9377.
9. **Gahmberg, N., E. Kuismanen, S. Keränen, and R. F. Pettersson.** 1986. Uukuniemi virus glycoproteins accumulate in and cause morphological changes of the Golgi complex in the absence of virus maturation. J. Virol. **57**:899–906.
10. **Gubler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene **25**:263–269.
11. **Henderson, L. E., S. Oroszlan, and W. Konigsberg.** 1979. A micromethod for complete removal of dodecyl sulfate from proteins by ion-pair extraction. Anal. Biochem. **93**:153–157.
12. **Hewlett, M. J., R. F. Pettersson, and D. Baltimore.** 1977. Circular forms of Uukuniemi virion RNA. An electron microscopic study. J. Virol. **21**:1085–1093.
13. **Ihara, T., H. Akashi, and D. H. L. Bishop.** 1984. Novel coding strategy (ambisense genomic RNA) revealed by sequence analyses of Punta Toro phlebovirus S RNA. Virology **136**:293–306.
14. **Kuismanen, E., B. Bång, M. Hurme, and R. F. Pettersson.** 1984. Uukuniemi virus maturation: immunofluorescence microscopy with monoclonal glycoprotein-specific antibodies. J. Virol. **51**: 137–146.
15. **Kuismanen, E., K. Hedman, J. Saraste, and R. F. Pettersson.** 1982. Uukuniemi virus maturation: accumulation of virus particles and viral antigens in the Golgi complex. Mol. Cell. Biol. **2**:1444–1458.
16. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
17. **Maizel, J. V., and R. P. Lenk.** 1981. Enhanced graphic matrix analyses of nucleic acid and protein sequences. Proc. Natl. Acad. Sci. USA **78**:7665–7669.
18. **Marriott, A. C., V. K. Ward, and P. A. Nuttall.** 1989. The S RNA segment of sandfly fever Sicilian virus: evidence for an ambisense genome. Virology **169**:341–345.
19. **Needleman, S. B., and C. D. Wunsch.** 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. **48**:443–453.
20. **Pettersson, R. F., N. Gahmberg, E. Kuismanen, L. Kääriäinen, R. Rönnholm, and J. Saraste.** 1988. Bunyavirus membrane glycoproteins as models for Golgi-specific proteins. Mod. Cell Biol. **6**:65–96.
21. **Pettersson, R. F., M. J. Hewlett, D. Baltimore, and J. M. Coffin.** 1977. The genome of Uukuniemi virus consist of three unique

RNA segments. Cell **11**:51–64.
22. **Pettersson, R. F., and L. Kääriäinen.** 1973. The ribonucleic acids of Uukuniemi virus, a noncubical tick-borne arbovirus. Virology **56**:608–619.
23. **Pettersson, R. F., E. Kuismanen, R. Rönnholm, and I. Ulmanen.** 1985. mRNAs of Uukuniemi virus, a bunyavirus, p. 283–300. *In* Y. Becker (ed.), Viral messenger RNA. Martinus Nijhoff Publishing, Boston.
24. **Pettersson, R. F., and C.-H. von Bonsdorff.** 1975. Ribonucleoproteins of Uukuniemi virus are circular. J. Virol. **15**:386–393.
25. **Raju, R., and D. Kolakovsky.** 1989. The ends of La Crosse virus genome and antigenome RNAs within nucleocapsids are base paired. J. Virol. **63**:122–128.
26. **Ranki, M., and R. F. Pettersson.** 1975. Uukuniemi virus contains an RNA polymerase. J. Virol. **16**:1420–1425.
27. **Rönnholm, R., and R. F. Pettersson.** 1987. Complete nucleotide sequence of the M RNA segment of Uukuniemi virus encoding the membrane glycoproteins G1 and G2. Virology **160**:191–202.
28. **Salser, W.** 1977. Globin mRNA sequences: analysis of base pairing and evolutionary implications. Cold Spring Harbor Symp. Quant. Biol. **42**:985–1002.
29. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
30. **Schmaljohn, C. S., G. B. Jennings, J. Hay, and J. M. Dalrymple.** 1986. Coding strategy of the S genome segment of Hantaan virus. Virology **155**:633–643.
31. **Ulmanen, I., P. Seppälä, and R. F. Pettersson.** 1981. In vitro translation of Uukuniemi virus-specific RNAs: identification of a nonstructural protein and a precursor to the membrane glycoproteins. J. Virol. **37**:72–79.
32. **von Bonsdorff, C.-H., and R. F. Pettersson.** 1975. Surface structure of Uukuniemi virus. J. Virol. **16**:1296–1307.